607

Consistent variance-covariance matrix estimation

Jon Faust

http://e105.org/e607

October 4, 2016

▶ **Readings/viewings: Stock and Watson course**

- Stock and Watson gave an NBER mini-course on econometrics a few years ago.

  These are very good

- Slides and videos are up.

  go

  http://www.nber.org/minicourse%5F2008.html

- Lecture 9 is on heteroskedasticity and autocorrelation consistent standard errors (HAC)

▶ **Hansen's text**

- Hansen's text (linked on the website) is also good on this topic

  sec 6.4, 6.5, 6.6

- Green and Hamilton also cover much of this topic

▶ **DGP**

- Take the DGP:

  $$y_t = x_t'\beta + \varepsilon_t$$

  or in matrix form:

  $$Y = X\beta + \varepsilon$$

  bn $Y, \varepsilon(T \times 1), X(T \times K), \beta(K \times 1)$

- Note x_t is a column vector made from the t^{th} row of X.

- The OLS estimator is

$$\hat{\beta} = (X'X)^{-1}X'Y$$

  and it is useful to define,

$$\hat{\varepsilon} = Y - X\hat{\beta}$$

- Substituting for Y using the DGP:

$$(\hat{\beta} - \beta) = (X'X)^{-1}X'\varepsilon$$

▶ $(\hat{\beta} - \beta)$ **is a func. of sample means**

$$\begin{aligned}(\hat{\beta} - \beta) &= (T^{-1}X'X)^{-1}T^{-1}X'\varepsilon \\ (\hat{\beta} - \beta) &= \bar{Q}^{-1}\bar{w}\end{aligned}$$

where

$$\begin{aligned}\bar{Q} &= T^{-1}\sum Q_t, \quad (K \times K) \\ \bar{w} &= T^{-1}\sum w_t, \quad (K \times 1) \\ Q_t &= x_t x_t' \\ w_t &= x_t \varepsilon_t\end{aligned}$$

- A1'. The xs are stochastic and satisfy

$$Ex_t x_t' = Q_t^e \quad (fullrank)$$

  and

$$T^{-1}\sum Q_t^e \to Q \quad (fullrank)$$

  and a WLLN applies to Q_t so that

$$\bar{Q} \to_p Q$$

- A2' (final). $E[\varepsilon_t|x_t] = 0$; $w_t = x_t\varepsilon_t$ are sufficiently well behaved that a WLLN and CLT apply:

$$\begin{aligned}\bar{w} &\to_p 0 \\ \sqrt{T}\bar{w} &\to_d N(0, \Omega)\end{aligned}$$

  for finite, full rank Ω.

- A3' The w_ts are sufficiently well behaved that we have a consistent estimator for $\hat{\Omega}$ for Ω.

▶ **And under these assumptions,**

- Under A1'–A3',

$$\sqrt{T}(\hat{\beta} - \beta) \rightarrow_d N(0, Q^{-1}\hat{\Omega}Q^{-1})$$

- This lecture discusses what assumptions will support consistent estimation of Ω and discusses actual estimators.

► **Note:**

- Given a consistent estimator, $\hat{\Omega}$ for Ω, we know that

$$\hat{V} = \bar{Q}^{-1}\Omega\bar{Q}^{-1}$$

will be consistent for the asymptotic variance-covariance matrix of $\hat{\beta}$.

Given that $\bar{Q}$ is consistent for Q.

- Up to now, we've said w_t is sufficiently well behaved for a CLT to apply to $\sqrt{T}\bar{w}$
- Now explicitly state that w_t has finite second moments variances, autocovariances, etc.
- Since we are not assuming the ws are cov. stat. these second moments can differ through time.
- Define

$$Ew_t w_s' \equiv \Sigma(t, s)$$

by finite second moments, we mean all of these are finite.

- Note that $\Sigma(t, s) = \Sigma(s, t)'$.
- But that these Σs can vary with t and s.

► **Some deep but trivial algebra**

- By definition, the exact variance of $\sqrt{T}\bar{w}$ is

$$\begin{aligned} \text{vcov}(\sqrt{T}\bar{w}) &= TE\bar{w}\bar{w}' \\ \Omega_T &= T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\Sigma(t, s) \end{aligned}$$

- Suppose we define $\bar{\Sigma}j$ to be the average of all the terms with $t < s$ and $j = |t - s|$:

$$\bar{\Sigma}(j) \equiv (T - j)^{-1}\sum_{t=1}^{T-j}\Sigma(t, t + j)$$

- We can now write Ω as:

$$\Omega = \bar{\Sigma}(0) + \sum_{j=1}^{T-1}\frac{T - j}{T}(\bar{\Sigma}(j) + \bar{\Sigma}(j)')$$

- Ponder this a bit, it is just regrouping and reorganizing the sum given above.

▶ **Crucial**

- This is just algebra on the definition of the variance in the general case:

$$\Omega = \bar{\Sigma}(0) + \sum_{j=1}^{T-1} \frac{T-j}{T}(\bar{\Sigma}(j) + \bar{\Sigma}(j)')$$

- Much of modern econometrics in this area arises essentially from ideas like: Hey, lets truncate that sum at a lower value than $T-1$; Hey, lets stick sample moments in for those population moments.
- This is trivial and could have been done, e.g., in 1930
- But much of the work was done during my career!?!
- What we were missing (mostly) was how to frame the problem, not the tools to solve it once properly framed.

▶ **Aside:: The scalar, cov. stat. case**

- If w is cov. stat. then for fixed j, $\Sigma(t, t+j)$ is the same for all t

  that is, the value of $\Sigma(t, s)$ depends only on $|t-s|$, second moments are constant.

- If w is a scalar then so is Σ
- So let's drop the bar on $\bar{\Sigma}(j)$ and move to lower case Σ in this scalar cov. stat. case.
- The formula is now:

$$\Omega = \sigma(0) + 2\sum_{j=1}^{T-1} \frac{T-j}{T}\sigma(j)$$

- The general formula above is a natural generalization of this for the vector and non-cov. stat. case.
- We will use this formula, which happens to be an expression for the spectral density of w at frequency zero, a good deal.

▶ **Back to main case**

- $$\Omega = \bar{\Sigma}(0) + \sum_{j=1}^{T-1} \frac{T-j}{T}(\bar{\Sigma}(j) + \bar{\Sigma}(j)')$$

- The theory underlying the CAN-WCEAVCM framework is largely a matter of making assumptions that simplify the form of Ω so that we can understand it, and then figuring out assumptions under which there is a consistent estimator.

- Much of this is trivial (once you have the idea): hey, stick sample moments in...

▶ **In this lecture**

- Explore consistent estimators for Ω under different assumptions
- And the nature of the assumptions we need for those estimators to be consistent.

▶ **Start with simplest case**

- To get oriented, let's start with the simplest case: ε_t is conditionally homoskedastic and not serially correlated:

$$\begin{aligned} E\varepsilon_t^2|x_t &= \sigma_\varepsilon^2 \\ E\varepsilon_t|x_t, x_{t-}, \varepsilon_{t-} &= 0 \end{aligned}$$

  where $t-$ means all dates prior to t

▶ **Simplest case**

- In this case, $\bar{\Sigma}(j) = 0$ for $j > 0$.
- Thus,

$$\begin{aligned} \Omega_T &= \bar{\Sigma}(0) \\ &\equiv T^{-1} \sum \Sigma(t,t) \\ &\equiv T^{-1} \sum E w_t w_t' \end{aligned}$$

- And since $w_t w_t' = x_t x_t' \varepsilon_t^2$, we have

$$\begin{aligned} E w_t w_t' &= E x_t x_t' E\varepsilon_t^2|x_t \\ &= Q_t \sigma_\varepsilon^2 \end{aligned}$$

- So,

$$\Omega = Q\sigma_\varepsilon^2$$

- That is, the OLS estimator.

▶ **The heteroskedasticity case: White's estimator**

▶ **White's estimator**

- Now allow conditional heteroskedasticity of ε and hence heterosked. of w.
- That is, $E[\varepsilon_t^2|x_t]$ varies

- But maintain the lack of serial correlation.
- We still have

$$\Omega_T = \bar{\Sigma}(0) = T^{-1} \sum E w_t w_t'$$

- The variance-covariance matrix is just the average of the variance-covariance matrices of the individual w_ts.
- But $Ew_t w_t'$ can vary arbitrarily (so far)
- White's idea. Hey, stick $w_t w_t'$ in for $Ew_t w_t'$:

$$\hat{\Omega_T} = T^{-1} \sum_{t=1}^{T} w_t w_t'$$

- So long as the sample mean of the $w_t w_t'$s satisfies a WLLN, then the sample mean of the $w_t w_t'$s will converge to the average of the $Ew_t w_t'$s.
- That is,

$$\hat{\Omega} - \Omega_T \to_p 0$$

- Intuition: we are using the sample average to estimate the average of the varying population moments

▶ **Aside:: WLLN with varying moments**

- Our standard WLLNs have $Ey_t = \mu$ and show $\bar{y} \to_p \mu$.
- It is fairly straightforward to generalize to $Ey_t = \mu_t$ and

$$\bar{y} - \bar{\mu} \to_p 0$$

  where $\bar{\mu} = T^{-1} \sum \mu_t$
- Simply take any process satisfying the original WLLN and add an arbitrary sequence $\{\delta_t\}$ to it so that $Ey_t = \mu + \delta_t$.

  The new sequence satisfies the re-stated WLLN because δ_t falls out of $\bar{y} - \bar{\mu}$.

▶ **Moving on**

- We now turn to the case when all the Σs are non-zero and vary

  Thus, we potentially have heterosked. and autocorrelation of the ws.

- Our desired estimator is known as a 'HAC' estimator for **H**eterosked. and **A**utocorrelation **C**onsistent.

▶ **The general formula**

- We go back to the general formula that holds in all cases:

$$\Omega_T = \bar{\Sigma}(0) + \sum_{t=1}^{T-1} \frac{T-j}{T}(\bar{\Sigma}(j) + \bar{\Sigma}(j)')$$

- How about we just follow the White-style logic above?
- That is, define

$$\hat{\Sigma}(j) = (T-j)^{-1} \sum_{t=1}^{T-j} x_t x_{t+j}'$$

  and stick the hat versions of the Σs in the Ω_T formula.
- This is essentially what we do

▶ **But, 2 practical problems**

- First, our Ω_T estimator sums the $\hat{\Sigma}(j)$s up to $j = T-1$.
- Notice we just have one observation to use in estimating $\hat{\Sigma}(T-1)$

  This is a covariance at lag $T-1$ and we have just one pair of observations this far apart in a sample of size T.
- We can't get a consistent estimator using 1 observation to estimate a quantity.

  Or 2 observations to estimate $\hat{\Sigma}(T-2)$ and so forth.
- Second, we know that Ω_T is positive definite, and in many of our applications, we'll need our $\hat{\Omega}$ to be positive definite as well.
- Nothing guarantees that if we put the sample $\hat{\Sigma}$s into the general formula, we will get a positive definite matrix out.

▶ **Solving the two problems**

▶ **Truncating the sum**

- One solution to the first problem is to truncate the big sum in the general formula

  That is, only include $\hat{\Sigma}$s that have enough observations to be fairly precisely estimated.
- but for consistency, as T rises, the precision of the estimator must rise.
- So for the consistency proof, we truncate the sum as a function of T

▶ **Some intuition for why this works**

- Remember that for sufficiently large J $\bar{\Sigma}(j)$ must small

  dependence must decay.

- Thus, if we truncate at some large value, the items we leave out will be small.

▶ **Truncation: a bit more intuition**

- Suppose we truncate the sum at $J = T^{1/4}$
- The last included $\hat{\Sigma}(j)$ included will be at lag $int(T^{1/4})$.
- Since we have T observations, there are approximately $T^{3/4} = \frac{T}{T^{1/4}}$ observations that are $T^{1/4}$ apart in the sample.
- Thus, as T grows, the included $\hat{\Sigma}$ that is based on the fewest observations will have a growing number of observations.
- And because the truncation lag is growing, we are leaving out more and more trivial things.

▶ **Aside:: Truncation as a clever estimator**

- We can think of truncating the sum in Ω_T as 'leaving out' some $\bar{\Sigma}$s.
- But we can instead think of this as using a clever estimator of these $\bar{\Sigma}$s:
- Specifically, we are using the estimator 0 to estimate $\bar{\Sigma}(j)$ for large j.
- It turns out that the famous estimator, 0, is often very useful.

▶ **Aside::**

- This estimator is biased, but if the thing being estimated is small, the bias is small.
- And this estimator hits the overall lower bound on variance: zero.
- What you lose on bias you may make up for with low variance.

▶ **Aside::**

- This is the first example in this class of a very important result you should always keep in the forefront of your mind in applied macroeconometrics:
- Pragmatic but false restrictions can increase the relevant sense of precision by decreasing variance by more than they increase bias.

▶ **Finally, the positive definite problem**

- Newey and West created a very famous HAC estimator.
- They started with truncating the sum and sticking sample moments in for population moments, but then figured out how to elegantly guarantee that the implied $\hat{\Omega}$ would be positive definite.
- The key is the weights in the sum.
- Notice that the weight applied to $\hat{\Sigma}(j)$ in the general formula is $(T - j)/T$.

- For fixed j this goes to 1 with T.
- It turns out that we can attain consistency using pretty much any weights that have this property of converging to 1.
- Newey and West characterized the entire class of weights that converge to 1 and that also guarantee positive definiteness

  under a particular choice of $\hat{\Sigma}(j)$.
- They proposed using $\frac{J+1-j}{J+1}$ where J is, as above, where we truncate the sum.
- Thus, the Newey-West estimator:

$$\hat{\Omega} = \hat{\Sigma}(0) + \sum_{t=1}^{J} \frac{J+1-j}{J+1}(\hat{\Sigma}(j) + \hat{\Sigma}(j)')$$

- Now we have an estimator, but have not been very specific about just what limits on good behavior of the w are sufficient for this estimator to be consistent
- I'll mention the major assumptions as illustrative of the sort of assumptions needed in this branch of theory.

► **Newey-West sense of suffic. well behaved**

- If you read Newey-West, you'll see that they are dealing with a more general GMM case so some additional notation and restrictions are needed.
- The key restrictions I want to emphasize is that w_t must have slightly more than 4 bounded moments:

$$E|w_t|^{4+r} < D < \infty$$

  for all t, and dependence in the data satisfy a mixing condition such that dependence decays at a rate defined in terms of r (same r as in the moment condition)
- The larger is r, the more finite moments and the slower the dependence can decay.
- As mentioned before, these conditions are not easy to check in practice

► **Wrapping up: WCEAVCM**

- In all cases, we have that the asymptotic variance-covariance matrix of OLS is $Q^{-1}\Omega Q$ and

$$\Omega = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \Sigma(t,s)$$

- We need a consistent estimator.

- Different assumptions put different structure on that double sum above and give rise to different estimators
- In the general case, we usually use something like Newey-West.
- The next notes deal with more pragmatic issues in estimating asymptotic variance-covariance matrices.